The future environmental impact of irrigated agriculture is grossly underestimated

Supplementary Material

Arnald Puy,[1,2*] Samuele Lo Piano,[3] Andrea Saltelli[2,3]

[1]Department of Ecology and Evolutionary Biology, Princeton University ,
M31 Guyot Hall, Princeton, New Jersey 08544, USA
[2]SVT, University of Bergen, Parkveien 9, PB 7805, 5020 Bergen, Norway
[3]Universitat Oberta de Catalunya, Edifici 22@, 08018 Barcelona, Spain

*To whom correspondence should be addressed; E-mail: apuy@princeton.edu

1 The model

Our model predicts future irrigated areas as a function of population, cropland and water available for irrigation. Population is a major force conditioning the extension of irrigated agriculture and agrarian systems at large (*30*, *31*), widely used as the main parameter to project irrigation-related variables in many modelling exercises (*32–34*). The relation between population and irrigated agriculture was first conceptualized by Boserup (*30*), who linked population density with intensive agrarian systems [see (*35*) for a summary of Boserup's influence]. At the continental level, however, population density does not explain the changes in the area under irrigation as much as population size (*36*) (Figs. S1–S2). Although other parameters might also play a key role in defining the size of irrigated agriculture (i.e. trans-oceanic food demands and trade networks), the available data and literature suggests that their effective influence might be

limited or well summarised by population size, a parameter that encapsulates and mirrors many complex properties and dynamics of social-ecological systems (*37–39*).

We described the relation between irrigated areas and population with the following equation:

$$Y = \alpha N^{\beta} \tag{1}$$

where Y and N are respectively the irrigated area and the population size, α is a constant and β is the so-called scaling exponent. If we log-transform Equation 1, we obtain

$$\log(Y) = \log(\alpha) + \beta \log(N) \tag{2}$$

where α and β can respectively be related to the intercept and the slope of the line of best fit.

We then defined a population model that allows for unexpected, occasional fast/slow population bursts and drops. Super-exponential population growth has been the norm for most of the known human history, and only very recently there has been a population slowdown (*48*, *49*). The United Nations projects this deceleration in the population growth rates up to 2100 (*49*). However, based on the systemic unpredictability characterizing the pace of population dynamics, and the fact that the United Nations has been systematically revising upwards their previsions since 2013, we considered this slowdown non-definitive, and assumed that some fluctuations in the population growth/decline speed should still be expected.

In our model, population growth was defined as

$$dN/dt = rN^{\gamma} \tag{3}$$

where r and N respectively reflect the growth rate and the population size, and γ is an ex-

ponent that controls the speed of population growth/decline. When $\gamma = 1$, population increases ($r > 0$) or decreases ($r < 0$) exponentially. When $\gamma > 1$, population increases ($r > 0$) or decreases ($r < 0$) faster than exponentially; when $\gamma < 1$, population increases ($r > 0$) or decreases ($r < 0$) slower than exponentially.

By re-arranging Equation 3, we obtained

$$\begin{aligned} dN/N^{\gamma} &= rdt \\ \frac{1}{1-\gamma}\left[N^{(1-\gamma)} - N_0^{(1-\gamma)}\right] &= rt \\ N &= \left[N_0^{(1-\gamma)} + rt(1-\gamma)\right]^{\frac{1}{1-\gamma}} \end{aligned} \tag{4}$$

and therefore,

$$N^{\beta} = \left[N_0^{(1-\gamma)} + rt(1-\gamma)\right]^{\frac{\beta}{1-\gamma}} \tag{5}$$

We then solved for dY/dt, and obtained

$$\begin{aligned} Y &= Y_0 + \alpha d(\left[N_0^{(1-\gamma)} + rt(1-\gamma)\right]^{\frac{\beta}{1-\gamma}}/dt \\ &= Y_0 + \alpha r(1-\gamma)\frac{\beta}{(1-\gamma)}\left[N_0^{(1-\gamma)} + rt(1-\gamma)\right]^{\left[\frac{\beta}{(1-\gamma)}-1\right]} \end{aligned} \tag{6}$$

where Y_0 is the irrigated area baseline value.

By simplifying Equation 6, it reduces to:

$$Y = Y_0 + \alpha\beta r\left[N_0^{(1-\gamma)} + rt(1-\gamma)\right]^{\left[\frac{\beta+\gamma-1}{1-\gamma}\right]} \tag{7}$$

In this model, α and β are correlated. Since we plan a sensitivity analysis as part of our modelling exercise (see Section 5), and because sensitivity analysis on correlated inputs is more cumbersome to perform and interpret than for uncorrelated ones (*50*), we re-arranged α as a function of β, as

$$\log(\alpha) = a + b\beta + \varepsilon \tag{8}$$

where a is the intercept, b the slope, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and σ the residual standard error of regressing α against β. The parameter ε thus describes the uncertainty linked to the prediction of α by β. By back-transforming α from its logarithm and plugging it into Equation 7, we obtained the final model:

$$Y = Y_0 + 10^{(a+b\beta+\varepsilon)}\beta r\left[N_0^{(1-\gamma)} + rt(1-\gamma)\right]^{\left[\frac{\beta+\gamma-1}{1-\gamma}\right]} \tag{9}$$

Equation 9 implies that the growth rate of irrigated areas is open-ended. This concurs with recent evidence suggesting that continental irrigated areas, on average, grow super-exponentially for a given increase in population (*36*). However, it can be argued that the growth of irrigated areas is constrained by the availability of cropland and water for irrigation. Such premise draws from the ecological concept of "carrying capacity" the idea of a static boundary that limits how much quantity of a given thing can exist in a given environment, irrigated agriculture in this case. The adoption of new approaches to save land and water, such as vertical agriculture (*51*) or seawater desalination (*52*), has shown this boundary to be dynamic. Yet we acknowledge the implausibility of 1) irrigated areas increasing from the current $\sim$6% to beyond 100% of the cultivable area in 30 years time, and 2) irrigated areas consuming all freshwater available, as this will leave no water for industrial or domestic use. Setting the available cropland and water for irrigation as upper boundaries for the growth of irrigated areas is therefore realistic, and has the advantage of preventing the model from becoming too cumbersome, thus preserving its empirically-grounded open-ended spirit and ease of interpretation.

In our model, we limited the growth of irrigated areas by the maximum cropland available as follows:

$$f(x) = \begin{cases} Y & Y < K \\ K & \text{otherwise} \end{cases} \tag{10}$$

where $f(x)$ is our new model, Y comes from Equation 9 and K is the maximum cropland available.

Regarding the water availability constraint, the demand of water for irrigation is conditioned by multiple factors [i.e. soil type, crop requirements, growing season, weather, irrigation technology/efficiency or the irrigated area size, among others (*53*, *54*)]. However, the latter describes a significant portion of the variability in the water withdrawn/required for irrigation (Fig. S3). Usually, and for a given irrigated area, the water withdrawn for irrigation is larger than the water required for irrigation due to water losses caused by conveyance, channel inefficiencies and evaporation. Aiming at keeping our model simple and intelligible, and following the linear trend shown in Fig. S3, we modelled the relation between water required for irrigation and irrigated areas as follows

$$\log(w) = \phi + \delta \log(Y) \tag{11}$$

where w is the volume of water required to irrigate the area Y, ϕ the intercept and δ the slope of the line of best fit.

We then limited the growth of irrigated areas by the maximum volume of water available for irrigation, which we made a fraction of the total water available: if the water required to irrigate the area Y is smaller than the total water that can be allocated to irrigation, we keep Y as the model output. If it is larger, we keep as the model output the largest irrigated area that can be irrigated with the total water available for irrigation:

$$f(x) = \begin{cases} Y & w < W_a \eta \\ (\frac{W_a \eta}{10^{\phi}})^{(1/\delta)} & \text{otherwise} \end{cases} \tag{12}$$

where $f(x)$ is our new model, Y comes from Equation 9, W_a is the total water available and η is the proportion of W_a that is allocated to irrigation.

2 The triggers

We made β a function of four triggers, as

$$\beta = f(X_1, X_2, X_3, X_4) \tag{13}$$

and (ϕ, δ) a function of three triggers, as

$$(\phi, \delta) = f(W_1, W_3, W_4) \tag{14}$$

Triggers are random parameters that allow to explore the uncertainty in the model design or in the calculation of a model parameter (see Section 3.2 for a detailed explanation of how the triggers work in our model). Their rationale are as follows:

- Triggers X_1 and W_1: There are currently six different data sets on global irrigated areas (*41–46*). The irrigated areas documented by these datasets diverge considerably in their size (Fig. S4), sometimes by more than two orders of magnitude at the national level (Figs. S5–S6). These different estimations introduce uncertainty in the value of Y in Equations 2 and 11, and in consequence, in the calculation of β and (ϕ, δ). Triggers X_1 and W_1 were thus used to select which of the six irrigated area data sets to use for the calculation of β and (ϕ, δ).

- Trigger X_2: β in Equation 2 can be computed via Ordinary Least Squares (OLS) or Standard Major Axis (SMA) regressions. SMA differs from OLS in that, instead of allocating all equation error to Y, it splits it into N and Y. This makes SMA unfit as a way to predict Y from N, but more appropriate when the aim is to summarize the bivariate scatter

between Y and N, i.e. when the relation between Y and N is symmetric and it is unclear which variable is the driver and which the response (*55*, *56*).

Predicting the extension of irrigation from population values requires assuming that population drives the behaviour of irrigated areas, and that irrigated areas do not influence population back. This assumption is an implicit by-product of the Boserup model (*30*), in which population had the status of an independent variable driving the adoption of intensive agriculture systems. Although this view has been adopted by many scholars, others have proposed that population increase/decrease can both lead and result from the intensification/desintensification of production, assuming that both parameters are inextricably intertwined (*57*). This latter position is more consistent with the characteristics of a SMA regression. Trigger X_2 was therefore used to decide whether OLS or SMA regressions should be used to compute α and β.

This rationale does not apply to the computation of (ϕ, δ) in Equation 11. The relation between irrigated areas and the water required for irrigation is asymmetric — the former (which is the driver) defines and conditions the latter (which is the response). In this context, OLS is the appropriate choice.

- Triggers X_3 and W_3. In the presence of outliers, OLS and SMA might provide different estimations of the line of best fit depending on whether robust or non-robust methods are used. We confirmed the presence of bivariate outliers by means of Mahalanobis classic and robust distances in both the combination of values for irrigated area / population (i.e. thus affecting the computation of α and β) and irrigated area / water required for irrigation (i.e. thus affecting the computation of ϕ and δ) (Figs. S7–S8) (*58*). Triggers X_3 and W_3 were thus used to select whether to apply robust or non-robust methods of estimation for β and (ϕ, δ) respectively.

- Triggers X_4 and W_4. Even if the aforementioned triggers account for the uncertainty in the computation of β and (ϕ, δ), there still would be uncertainty as regards to their true value. This can be approached by bootstrapping β (ϕ, δ) n times in each of the 24 (12) possible scenarios created by the triggers, thus creating 24 (12) different vectors of β (ϕ, δ) with length n per continent [see Section 3.2]. Triggers X_4 and W_4 were therefore used to decide which β and (ϕ, δ) values should respectively be extracted from the vector and used in the computation of the model output.

3 Estimation of the model parameters

We designed, coded and ran the model and the uncertainty/sensitivity analysis (see Section 5) both in the R environment (*59*) and in Python (*60*). The code and all data sets required to replicate our results are available from the authors here (R) and here (Python).

3.1 Irrigated area baseline values (Y_0)

We used the data compiled by Meier et al. (*43*) to characterize the uncertainty in the current extension of irrigation. Meier et al. collected previously published data sets, remote sensing-based vegetation activity and ancillary information to produce a map of global irrigated areas between 1999-2012. They also compared the resulting map with the irrigated areas previously reported by Thenkabail et al. (*46*), Siebert et al. (*45*), Salmon et al. (*44*), FAO's Aquastat (*41*) and FAOSTAT (*42*), which respectively reflected the extension under irrigation at the end of the last millennium, between 2000-2008, in the year 2005 and between 1999-2012.

We thus described irrigated area baseline values as $Y_{0,i} \sim \mathcal{U}(a_i, b_i)$ (continuous), with a_i and b_i being respectively the minimum and maximum irrigated area for continent i according to the range of values covered by the six different data sets just mentioned. We assumed that any value comprised between a_i and b_i held the same probability of reflecting the true extension

of irrigation for continent i. We also considered that each single year between 1999-2012 held the same probability of being the baseline year for the projection of irrigated areas to 2050 (see Section 3.3).

3.2 Growth rate between population and irrigated area (β), irrigated area and water required for irrigation (ϕ, δ), and noise (ε)

We computed β and (ϕ, δ) for each combination of triggers (X_1, X_2, X_3) and (W_1, W_3) respectively.

Firstly, we followed Equations 2–11 and regressed the irrigated areas documented by each of the six available data sets against population and irrigation water requirement values. Data on population and water requirements for irrigation were respectively retrieved from the United Nations population division (*49*) and Aquastat (*41*). We estimated β by bootstrapping ($n = 10.000$) the coefficients obtained through OLS and SMA regressions, both robust and non-robust. For ϕ and δ, we relied on the same approach but used OLS only. We followed the algorithm by Salibian-Barrera (*61*) to compute robust OLS, and used the robust Huber's M estimator (*62*) for robust SMA. We then created a lookup table for β (ϕ, δ) with 24 (12) slots per continent, each including a vector of 10.000 bootstrapped β (ϕ, δ) samples computed according to a specific combination of (X_1, X_2, X_3) (W_1, W_3). The samples were sorted in ascending order, aiming at linking high values of X_4 (W_4) with high β (δ) values during the execution of the model (see Section 2).

Given the fact that $\beta = f(X_1, X_2, X_3, X_4)$ and $(\phi, \delta = f(W_1, W_3, W_4)$, β and (ϕ, δ) were intermediate parameters and therefore we did not include them in the sensitivity analysis stage.

Finally, we pooled all the bootstrapped samples from all the slots and followed Equation 8 to predict α from β and obtain values for a, b and ε.

3.3 Time (t)

We described time with a discrete uniform distribution, as $t \sim \mathcal{U}(a, b)$, with $a = 2050 - 2012 = 38$, and $b = 2050 - 1999 = 51$. The consideration of t as an uncertain parameter derives from the fact that the available irrigated area data sets reflect the extension of irrigation for an uncertain time frame (1999-2012) (see Section 3.1).

3.4 Population baseline values (N)

We used the United Nations population division (*47*) to retrieve yearly population data between 1999-2012. The value of N to be used in the model was defined by the baseline year and therefore by the parameter t. This made N deterministically dependent upon t and, in consequence, although we used it in the computation of the model output, we did not retain N for the sensitivity analysis.

3.5 Population growth rates (r)

We used the population growth rates published by the United Nations Population Division (*47*). Since we considered any year between 1999-2012 as a potential departure point for the projection of irrigated areas to 2050, we retrieved the population growth rates from 2000-2015 and the population growth rates estimated for 2015-2050. We pooled the data at the continental level and assessed the distributions by means of histograms. We then fitted possible distributions to the data, and selected the best fit by means of Q-Q plots, CDF plots, P-P plots and Bayesian Information Criteria. Following the results, we described population growth rates as $r \sim Wei(k, \gamma)$ for Africa, Asia and Europe, with k being the shape and γ the scale of the distribution, and as $r \sim \mathcal{N}(\mu, \sigma)$ for the Americas.

3.6 Pace of population growth/decline (γ)

We described the pace of population growth/decline as $\gamma \sim \mathcal{N}(1, 0.02)$. This allowed the population growth rate to eventually fluctuate from a dominant exponential to an occasional slightly sub and super-exponential population growth regime. We used a standard deviation of 0.02 in order to be conservative and enclose this fluctuation within reasonable bounds. Our conceptualization thus overlooks the possibility of unexpected catastrophic population drops or booms due to extreme or unlikely events such as global wars, pandemics or large-scale policies encouraging or forcing higher fertility rates.

3.7 Cropland available (K)

We used the study by Zhang and Cai (*63*) to characterize the distribution of K. Zhang and Cai (*63*) calculated both the net (i.e. cropland available minus human settlement areas and protected land) and the gross (i.e. including these areas) cropland available in 2050 for Africa, China, India, Europe, Russia, South America and the US. We constructed the range of possible values for the Americas by adding the minimum and maximum extension for South America and the US; and the range for Asia, by adding the minimum and maximum extension for China and India. The values for Russia were added to those of Europe. We described the maximum cropland available as $K_i \sim \mathcal{U}(a_i, b_i)$, with a_i and b_i being respectively the minimum net and maximum gross cropland available in 2050 for continent i.

3.8 Water available (W_a)

To characterize the water available in 2050, we used FAO's internal renewable water resources (IRWR) values in 2014 plus an uncertainty range of 20%, set on the basis of the communication of one of the authors (S.L.P) with an Aquastat/FAOSTAT responsible (Francesco Tubiello). IRWR is defined as the "long-term average annual flow of rivers and recharge of aquifers

generated from endogenous precipitation" (*41*). Since Aquastat computes IRWR at the country level, we aggregated the data to obtain a reasonable estimate of the total water currently available for agriculture, industrial and domestic purposes in each continent. We added the abovementioned uncertainty range over the final aggregated value and described W_a in 2050 as $W_a \sim \mathcal{U}(3142, 4714)$ for Africa, $W_a \sim \mathcal{U}(8851, 13277)$ for Asia, $W_a \sim \mathcal{U}(15629, 23553)$ for the Americas, and $W_a \sim \mathcal{U}(5261, 7891)$ for Europe (km^3/year).

3.9 Proportion of water available that can be allocated to irrigation (η)

Aquastat (*41*) has produced some estimates of the percentage of total water available currently allocated to agriculture, industries and municipalities (link here). There are also some studies that have attempted to project these figures up to 2100 (*64–66*). Since a proper treatment of the uncertainties derived from the prediction of water allocation across societal uses is beyond the scope of our study, we adopted a different approach: we characterized η using the globally-adopted Withdrawal-to-Availability ratio (WTA) index (*67*). According to the WTA, countries are water-stressed when annual withdrawals for agriculture, industries and domestic purposes are between 20–40% of the IRWR, and severely stressed when they are >40% (*68*). We considered that a severe to a critical pressure on water resources by irrigated agriculture only is a likely scenario in 2050, and defined η as $\eta \sim \mathcal{U}(0.2, 0.5)$.

4 Creation of the sample matrix

We generated a $(2^{17}, 2k)$ matrix for each continent, with k being the number of model parameters and triggers $(Y_0, X_1, ...)$, using Sobol' quasi-random number sequences (*69*). The first k-matrix was labelled **A** and the remaining k-matrix, **B**. The columns of the **A** and **B** matrices were linked to a specific parameter or trigger and transformed into their appropriate probability distribution.

The columns representing the triggers were transformed as follows: X_1 and W_1 were described as $(X_1, W_1) \sim \mathcal{U}(1, 6)$ (discrete), each number indicating the irrigated area data set to be used to calculate β, i.e. 1=Aquastat (*41*), 2=FAOSTAT (*42*), 3=Siebert et al. (*45*), 4=Meier et al. (*43*), 5=Salmon et al. (*44*), and 6=Thenkabail et al. (*46*). The column for trigger X_2 was recoded as $X_2 \sim \mathcal{U}(1, 2)$ (discrete), where 1=OLS and 2=SMA. The columns for triggers X_3 and W_3 were recoded as $\sim \mathcal{U}(1, 2)$, where 1= use robust regression and 2=use non-robust regression. The columns for triggers X_4 and W_4 were recoded as $(X_4, W_4) \sim \mathcal{U}(1, 10^4)$ (discrete), where 10^4 was the number of bootstrapped β and (ϕ, δ) values in each slot of the lookup tables (see Section 3.2).

The values of β (ϕ, δ) for the v-th row were extracted from the bootstrapped samples according to the conditions simultaneously set by triggers $X_1^{(v)}$ ($W_1^{(v)}$), $X_2^{(v)}$, $X_3^{(v)}$ ($W_3^{(v)}$) and $X_4^{(v)}$ ($W_4^{(v)}$). We then ran the model row-wise in both the **A** and the **B** matrix, obtaining two vectors of model outputs, $Y_{\mathbf{A}} = f(\mathbf{A})$ and $Y_{\mathbf{B}} = f(\mathbf{B})$. The uncertainty analysis focused on both the **A** and **B** matrices.

5 Sensitivity analysis

We relied on Sobol' indices to decompose the model output variance into fractions that can be apportioned to single parameter/triggers or groups of parameters/triggers (*70*). For a model of the form $Y = f(X_1, X_2, ..., X_k)$, where Y is a scalar and $X_1, X_2, ..., X_k$ are independent parameters described by known probability distributions, we can measure how sensible Y is to a given parameter X_i with

$$V_i = V_{X_i}\left[E_{\mathbf{X}_{\sim i}}(Y|X_i)\right] \tag{15}$$

where $E_{\mathbf{X}_{\sim i}}(Y|X_i)$ is the expected value of Y calculated over all possible values of all parameters except the i-th, which is kept fixed. By dividing Equation 15 by the unconditional

model output variance V_Y, we obtain the first order sensitivity index for X_i, which describes the proportion of variance in the model output caused by X_i:

$$S_i = \frac{V_i}{V_Y} \tag{16}$$

We can then decompose the unconditional model output variance Y as the sum of conditional variances up to the k-th order:

$$V_Y = \sum_{i=1}^{k} V_i + \sum_{i} \sum_{i<j} V_{ij} + ... + V_{1,2,...,k} \tag{17}$$

where

$$\begin{aligned} V_{ij} =& V_{X_i,X_j}\left[E_{\mathbf{X}_{\sim i,j}}(Y|X_i, X_j)\right] \\ &- V_{X_i}\left[E_{\mathbf{X}_{\sim i}}(Y|X_i)\right] \\ &- V_{X_j}\left[E_{\mathbf{X}_{\sim j}}(Y|X_j)\right] \end{aligned} \tag{18}$$

From this, we can derive the second-order index S_{ij}, which explains the proportion of variance due to the interaction between X_i and X_j:

$$S_{ij} = \frac{V_{ij}}{V_Y} \tag{19}$$

and so on until order k. However, estimating all terms in Equation 17 is unattainable when k is large, as they result in $2^k - 1$. In this case, we can compute the total order index or S_{Ti}, which measures the proportion of variance due to the first-order effect of X_i jointly with its interactions with the other parameters (*71*):

$$S_{Ti} = \frac{E_{X_{\sim i}}\left[V_{X_i}(Y|X_{\sim i}\right]}{V_Y} \tag{20}$$

To calculate Sobol' first (S_i) and total (S_{Ti}) order effects, we designed k additional matrices ($\mathbf{A}_{\mathbf{B}}^{i}$), $i = 1, 2, ..., k$, where the k matrix is composed of all columns of the **A** matrix except the

i-th column, which is the i column of the **B** matrix. For $n = 2^{17}$, this approach led to a total number of model runs per continent of $n(k+2) = 2^{17}(15+2) = 2.22$M (*72*).

We first calculated the sample mean $\hat{f}_0$ of the model output as follows:

$$\hat{f}_0 = \frac{1}{2n}\sum_{v=1}^{n}\left(f(\mathbf{A})^{(v)} + f(\mathbf{B})^{(v)}\right) \tag{21}$$

where $f(\mathbf{A})^{(v)}$ [$f(\mathbf{B})^{(v)}$] means the model output Y after running the model in the v-th row of the **A** [**B**] matrix.

We then computed the unconditional sample variance $\hat{V}_y$ as follows:

$$\hat{V}_y = \frac{1}{2n-1}\sum_{v=1}^{n}\left[(f(\mathbf{A})^{(v)} - \hat{f}_0)^2 + (f(\mathbf{B})^{(v)} - \hat{f}_0)^2\right] \tag{22}$$

We computed S_i after Saltelli et al. (*72*) and S_{Ti} after Jansen (*73*):

$$S_i = \frac{\frac{1}{n}\sum_{v=1}^{n} f(\mathbf{B})^{(v)}\left[f(\mathbf{A}_{\mathbf{B}}^{i})^{(v)} - f(\mathbf{A})^{(v)}\right]}{\hat{V}_y} \tag{23}$$

$$S_{Ti} = \frac{\frac{1}{2n}\sum_{v=1}^{n}\left[f(\mathbf{A})^{(v)} - f(\mathbf{A}_{\mathbf{B}}^{i})^{(v)}\right]^2}{\hat{V}_y} \tag{24}$$

We bootstrapped both S_i and S_{Ti} 1000 times in order to obtain the 95% confidence intervals. To compute the extent of the approximation error, we followed Khorashadi Zadeh et al. (*74*).

References

30. E. Boserup, *The Conditions of Agricultural Growth* (George Allen 6 Unwin Ltd, London, 1965).

31. M. Barceló, presented at the El Agua en Zonas Áridas. Arqueología e Historia. Actas del I Coloquio de Historia y Medio Físico. vol. I, ed. by L. Cara Barrionuevo.

32. I. Shiklomanov, J. Rodda, *World Water Resources at the Beginning of the 21st Century* (Cambridge University Press, Cambridge, 2003).

33. Y. Shen, T. Oki, N. Utsumi, S. Kanae, N. Hanasaki, *Hydrological Sciences* **53**, 11–33 (Feb. 2008).

34. N. Alexandratos, J. Bruinsma, "World Agriculture Towards 2030/2050. The 2012 Revision", tech. rep. (FAO, Rome, 2012), p. 375.

35. B. L. Turner II, M. Fischer-Kowalski, *Proceedings of the National Academy of Sciences* **107**, 21963–21965 (2010).

36. A. Puy, *Ecological Applications* **28**, 1413–1419 (Sept. 2018).

37. L. M. A. Bettencourt, *Science* **340**, 1438–1441 (2013).

38. R. Muneepeerakul, M. R. Qubbaj, *Ecological Economics* **77**, 123–128 (2012).

39. R. C. Oka, M. Kissel, M. Golitko, S. G. Sheridan, N. C. Kim, A. Fuentes, *Proceedings of the National Academy of Sciences* **114**, E11101–E11110 (Dec. 2017).

40. J. Lincoln Simon, *Population and Development in Poor Countries: Selected Essays* (Princeton Legacy Library, Princeton, 2014).

41. FAO, *AQUASTAT website*, 2016.

42. FAO, *FAOSTAT database*, Rome, 2017.

43. J. Meier, F. Zabel, W. Mauser, *Hydrology and Earth System Sciences* **22**, 1119–1133 (2018).

44. J. Salmon, M. A. Friedl, S. Frolking, D. Wisser, E. M. Douglas, *International Journal of Applied Earth Observation and Geoinformation* **38**, 321–334 (2015).

45. S. Siebert, V. Henrich, K. Frenken, J. Burke, *Update of the digital global map of irrigation areas to version 5*, Rome, 2013.

46. P. S. Thenkabail, C. M. Biradar, P. Noojipady, V. Dheeravath, Y. Li, M. Velpuri, M. Gumma, O. R. P. Gangalakunta, H. Turral, X. Cai, J. Vithanage, M. Schull, R. Dutta, *International Journal of Remote Sensing* **30**, 3679–3733 (2009).

47. United Nations, *World population prospects. The 2017 revision*, 2017.

48. A. Johansen, D. Sornette, *Physica A: Statistical Mechanics and its Applications* **294**, 465–502 (May 2001).

49. United Nations (2017).

50. A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global Sensitivity Analysis. The Primer* (John Wiley & Sons, Ltd, Chichester, UK, Dec. 2007).

51. D. Despommier, *Journal fur Verbraucherschutz und Lebensmittelsicherheit* **6**, 233–236 (2011).

52. S. Burn, M. Hoang, D. Zarzo, F. Olewniak, E. Campos, B. Bolto, O. Barron, *Desalination* **364**, 2–16 (2015).

53. D. Wisser, S. Frolking, E. M. Douglas, B. M. Fekete, C. J. Vörösmarty, A. H. Schumann, *Geophysical Research Letters* **35**, 1–5 (2008).

54. V. Chaturvedi, M. Hejazi, J. Edmonds, L. Clarke, P. Kyle, E. Davies, M. Wise, *Mitigation and Adaptation Strategies for Global Change* **20**, 389–407 (2013).

55. R. J. Smith, *American Journal of Physical Anthropology* **140**, 476–486 (2009).

56. D. I. Warton, I. J. Wright, D. Falster, M. Westoby, *Biological Reviews* **81**, 259 (May 2006).

57. K. D. Morrison, *Journal of Archaeological Method and Theory* **1**, 111–159 (1994).

58. P. Filzmoser, R. G. Garrett, C. Reimann, *Computers and Geosciences* **31**, 579–587 (2005).

59. R Core Team, in (R Foundation for Statistical Computing, Vienna, 2018).

60. Phyton Software Foundation, *Python Language Reference*.

61. M. Salibián-Barrera, S. Van Aelst, G. Willems, *Statistical Methods and Applications* **17**, 41–71 (2008).

62. D. I. Warton, R. A. Duursma, D. Falster, S. Taskinen, *Methods in Ecology and Evolution* **3**, 257–259 (Apr. 2012).

63. X. Zhang, X. Cai, *Environmental Research Letters* **6**, 014014 (Jan. 2011).

64. J. Alcamo, M. Flörke, M. Märker, *Hydrological Sciences Journal* **52**, 247–275 (2007).

65. M. Hejazi, J. Edmonds, L. Clarke, P. Kyle, E. Davies, V. Chaturvedi, M. Wise, P. Patel, J. Eom, K. Calvin, R. Moss, S. Kim, *Technological Forecasting and Social Change* **81**, 205–226 (2014).

66. Y. Wada, M. Flörke, N. Hanasaki, S. Eisner, G. Fischer, S. Tramberend, Y. Satoh, M. T. Van Vliet, P. Yillia, C. Ringler, P. Burek, D. Wiberg, *Geoscientific Model Development* **9**, 175–222 (2016).

67. S. Damkjaer, R. Taylor, *Ambio* **46**, 513–531 (2017).

68. J. Alcamo, P. Döll, T. Henrichs, F. Kaspar, B. Lehner, T. Rösch, S. Siebert, *Hydrological Sciences Journal* **48**, 339–348 (2003).

69. P. Bratley, B. L. Fox, *ACM Transactions on Mathematical Software (TOMS)* **14**, 88–100 (1988).

70. I. Sobol', *Mathematics and Computers in Simulation* **55**, 271–280 (Feb. 2001).

71. T. Homma, A. Saltelli, *Reliability Engineering & System Safety* **52**, 1–17 (1996).

72. A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, *Computer Physics Communications* **181**, 259–270 (Feb. 2010).

73. M. Jansen, *Computer Physics Communications* **117**, 35–43 (Mar. 1999).

74. F. Khorashadi Zadeh, J. Nossent, F. Sarrazin, F. Pianosi, A. van Griensven, T. Wagener, W. Bauwens, *Environmental Modelling and Software* **91**, 210–222 (2017).

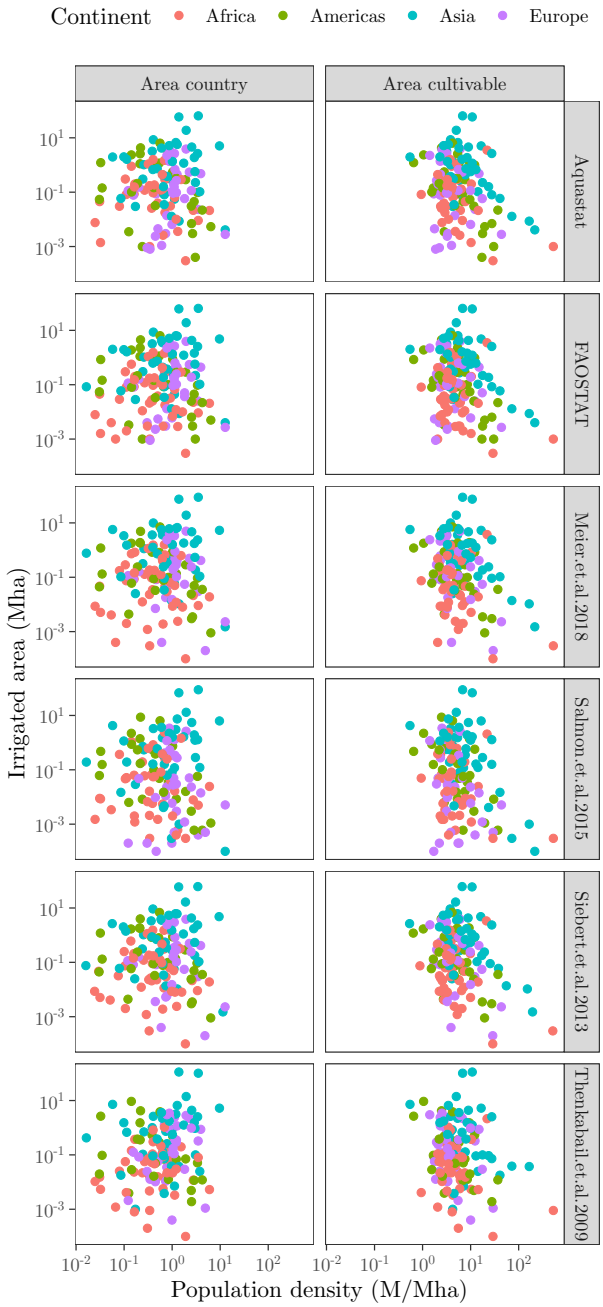

Figure S1: Scatter plots of measures of population density against irrigated areas. We have chosen two estimators of population density: "population density per unit of terrain", defined as *total population / total area of the country*, and "population density per unit of cultivable land", operationalized as *total population / total cultivable land*. The latter aims at ruling out uncultivable areas such as deserts, rocky mountains or water bodies, included in the calculation of the first estimator as part of the "total area of the country" variable (*40*). The data for the cultivable area and the country area have been retrieved from (*41*). The horizontal strip label indicates the data set used to plot the values for irrigated areas: Aquastat (*41*), FAOSTAT (*42*), Meier et al. (*43*), Salmon et al. (*44*), Siebert et al. (*45*) and Thenkabail et al. (*46*). All those data sets have been compiled and studied by Meier et al. (*43*) and reflect the extension under irrigation between 1999–2012.

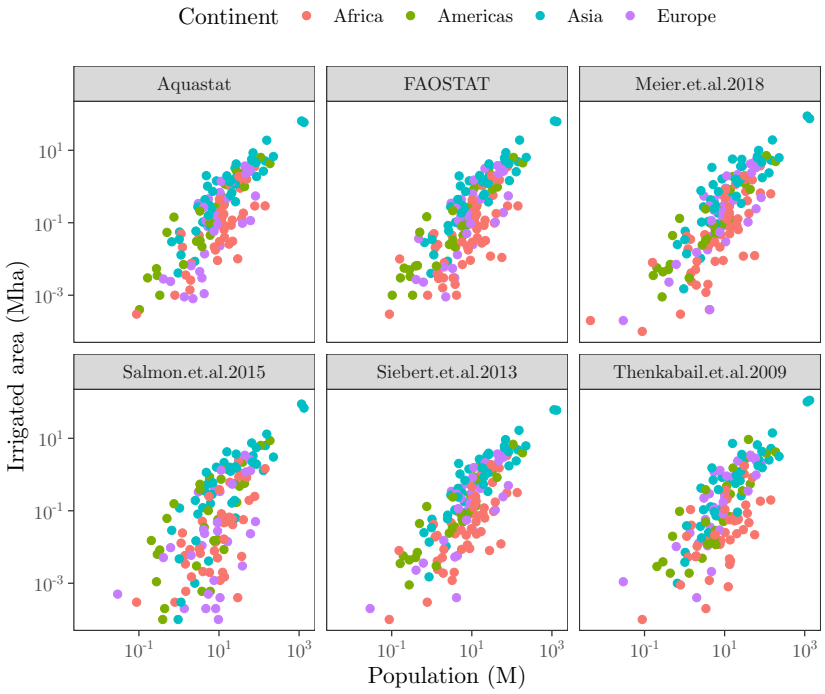

Figure S2: Scatter plots of irrigated areas against population size. For the description of the grey labels, see Fig. S1. The x axis reflects the mean population size (1999-2012), computed using data from the United Nations Population Division (*47*).

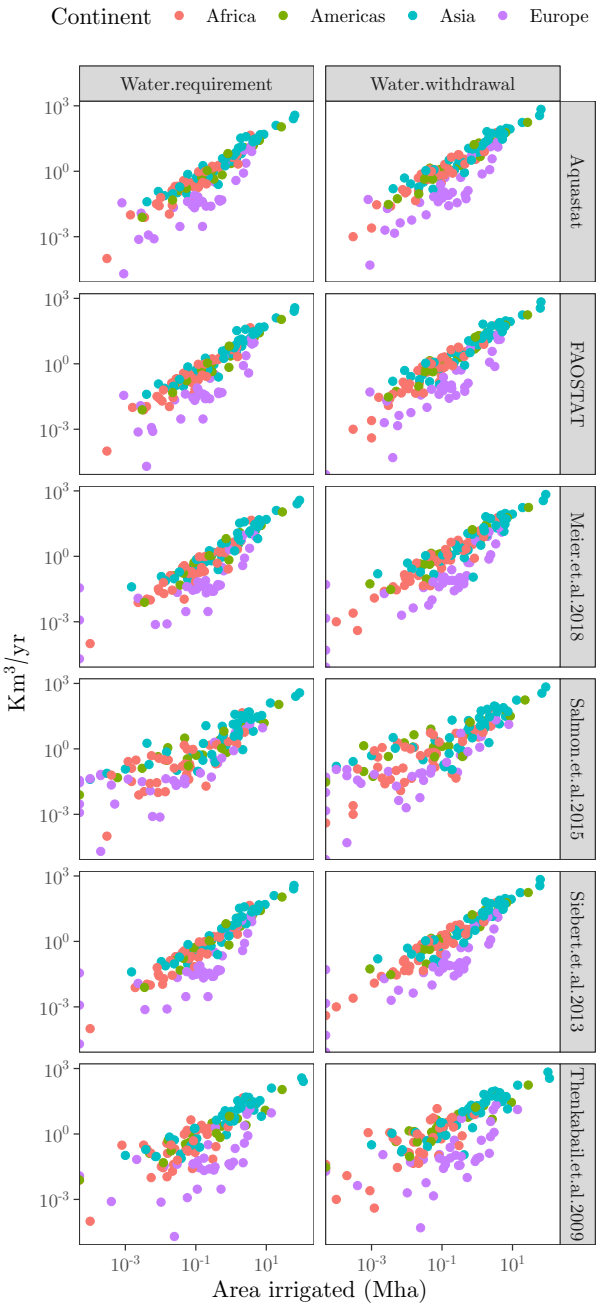

Figure S3: Scatter plots of irrigated areas against irrigation water requirement and irrigation water withdrawal. The former reflects the volume of water required for the crop to grow; the latter the total volume of water withdrawn for irrigation. The data water withdrawal and water requirement has been retrieved from Aquastat (*41*). For the description of the grey vertical labels, see Fig. S1.

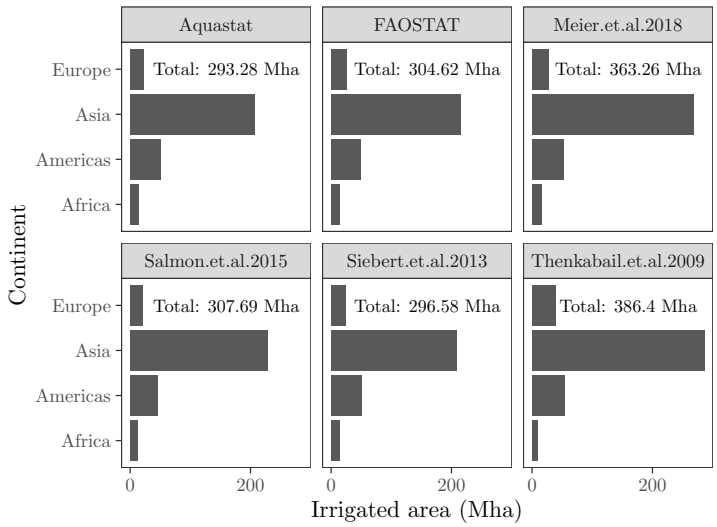

Figure S4: The extension of irrigation documented by different authors and institutions. Aquastat (*41*) and FAOSTAT (*42*) provide official values based on national surveys, census and statistics. Siebert et al. (*45*) merge FAOSTAT and Aquastat values with independent maps and remote sensing imagery. Salmon et al. (*44*) integrate national and subnational surveys with remote sensing and gridded climate data sets. Thenkabail et al. (*46*) rely on remote sensing, Google Earth, and ground control points. Meier et al. (*43*) downscale the map by Siebert et al. (*45*) and use multi-temporal normalized difference vegetation indexes with agricultural suitability data. The data was retrieved from Meier et al. (*43*)

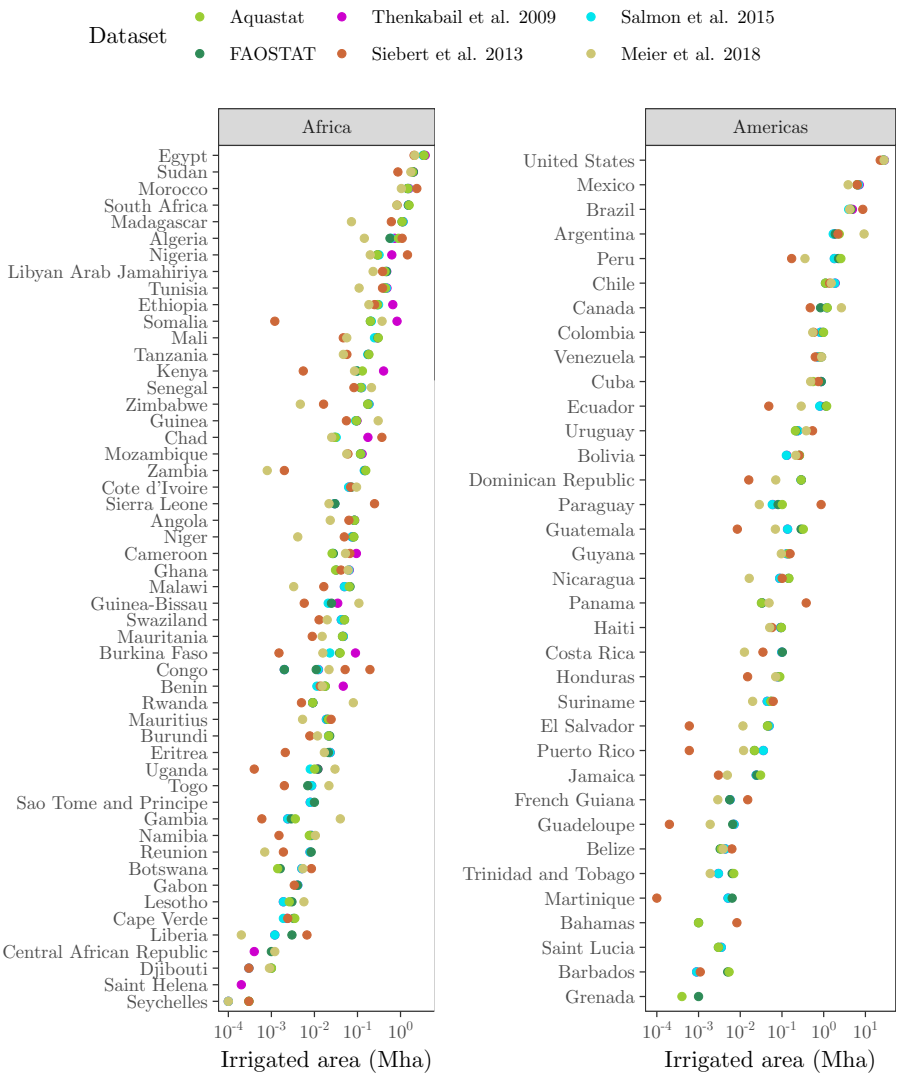

Figure S5: Extension of irrigation at the country level for Africa and the Americas. The data was retrieved from Meier et al. (*43*)

Figure S6: Extension of irrigation at the country level for Asia and Europe. The data was retrieved from Meier et al. (*43*)

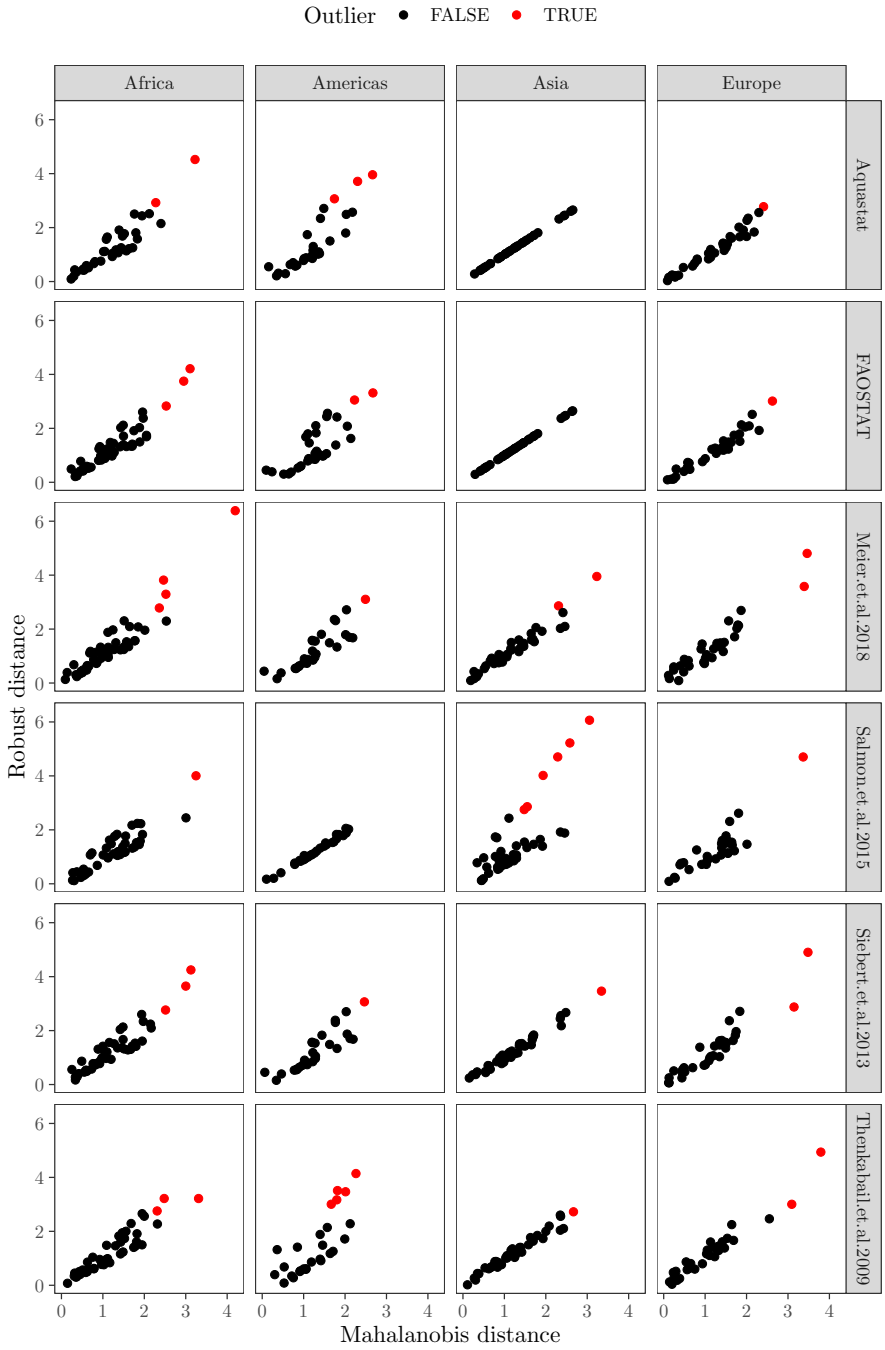

Figure S7: Scatter plot of Mahalanobis vs Robust distances for irrigated areas and population. The red dots represent countries whose combination of values for these two variables have been identified as multivariate outliers (*58*).

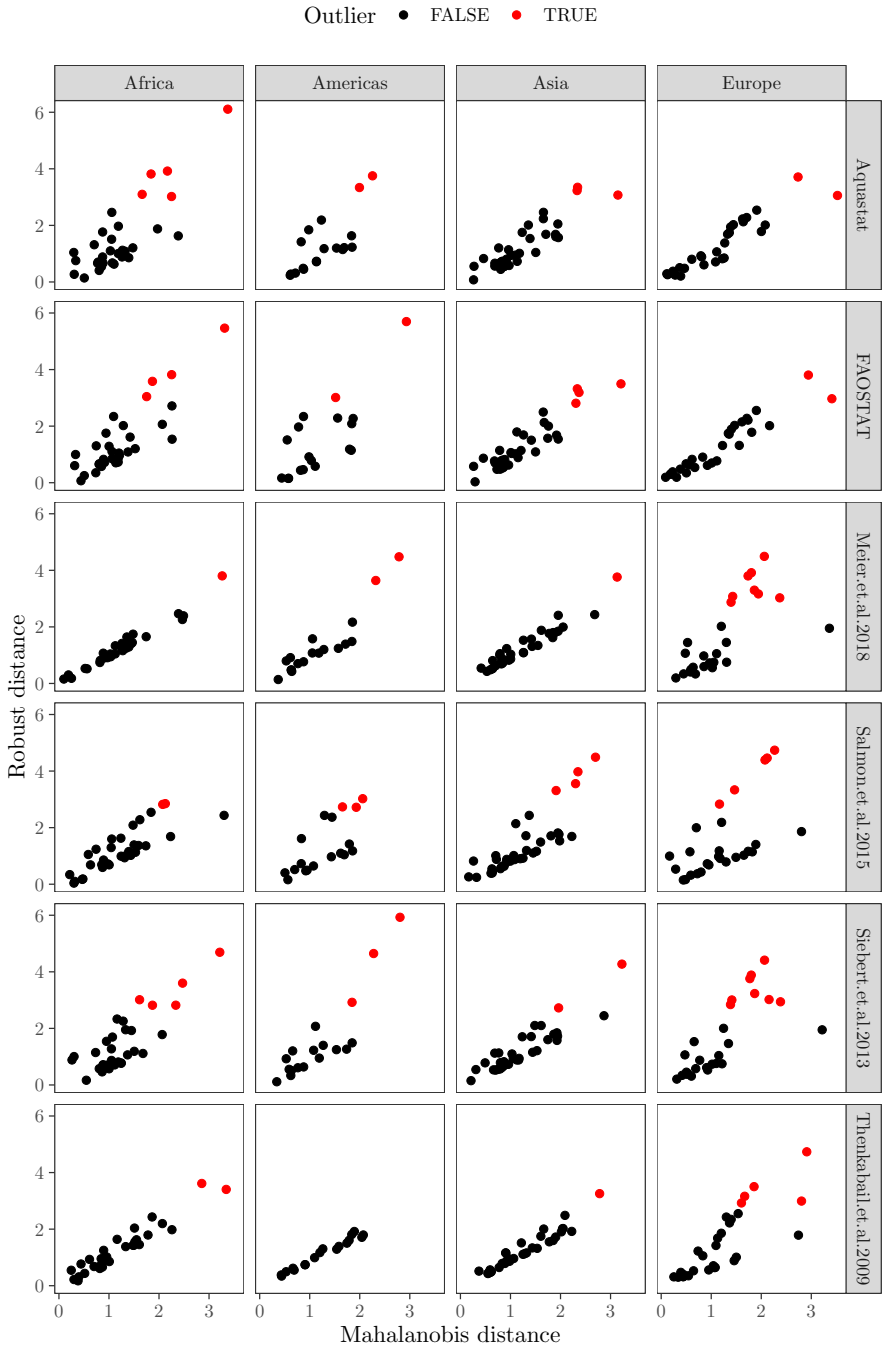

Figure S8: Scatter plot of Mahalanobis vs Robust distances for irrigated areas and water required for irrigation. The red dots represent countries whose combination of values for these two variables have been identified as multivariate outliers (*58*).